DOCUMENT RESUME

ED 239 219 CS 007 436

AUTHOR Shanahan, Timothy; Kamil, Michael L.
TITLE The Relationship of the Concurrent and Construct Validities of Cloze.
PUB DATE 83
NOTE 12p.; Paper presented at the Annual Meeting of the National Reading Conference (33rd, Austin, TX, November 29-December 3, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Age Differences; *Cloze Procedure; *Correlation; *Predictor Variables; *Reading Comprehension; *Reading Research; Reading Tests; Reliability; Research Methodology; *Test Validity
IDENTIFIERS *Bormuth (John R)

ABSTRACT

To investigate the causes of variation in cloze-comprehension test correlations, a reanalysis was conducted of the influential J.R. Bormuth study (1962), which reported correlations between nine cloze and nine comprehension tests administered to 50 subjects in grades 4, 5, and 6. Two separate renalayses were completed in the present study, the first using cloze-comprehension correlations as a dependent measure and the second using correlations of cloze with each of the Bormuth question catagories--vocabulary, factual recall, sequential order, cause and effect, inference, and author's purpose--at each grade level. Results indicated that in both reanalyses, variable reliability and grade level differences accounted for some of the variation in correlations. The proportion of comprehension test questions found to assess within-sentence information was found to be significant in the first regression analysis but not in the second. Results suggested that the limited construct validity of cloze as a measure of comprehension had a small but significant influence on the concurrent validity of cloze with measures of comprehension. Cloze predicted performance best when within-sentence comprehension dominated, but was less useful when students were expected to integrate information across sentence boundaries. (MM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED239219

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Timothy Shanahan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Timothy Shanahan
College of Education
Box 4348
University of Illinois
at Chicago
Chicago, IL 60680
(312) 996-4677 or 996-4508

Michael L. Kamil
College of Education
Box 4348
University of Illinois
at Chicago
Chicago, IL 60680
(312) 996-5626 or 996-4508

The Relationship of the Concurrent and Construct Validities of Cloze

Timothy Shanahan Michael L. Kamil

University of Illinois at Chicago

Cloze tests have traditionally been used for two purposes: to assess comprehension and to establish reading levels for instruction. However, recent cloze research (Kibby, 1981; Shanahan & Kamil, 1982; Shanahan & Kamil, in press; Shanahan, Kamil, & Tobin, 1982) has demonstrated that cloze has low construct validity as a test of reading comprehension. Specifically, cloze is insensitive to the integration of information across sentence boundaries. Thus, the use of cloze to measure comprehension should be viewed with suspicion.

Cloze is used to establish reading levels because it has been found to have high concurrent validity, not because of any claims about its construct validity. That is, cloze has been proposed as an attractive alternative to more

CS 007 436

expensive (in terms of time and effort) approaches to reading assessment because it has been found to be significantly correlated with such measures (Bickley, Ellington, & Bickley, 1970; McKenna & Robinson, 1980; Rankin, 1965).

The correlations of cloze scores with reading comprehension scores usually range from about .40 to .80, across a variety of tests, populations, etc. Cloze actually explains as little as 16% to as much as 65% of reading comprehension test score variance. Although these correlations are usually significant, this range of values suggests that cloze is more appropriate under certain text or measurement conditions than it is under others. The purpose of this paper is to identify sources of variation in a set of cloze-comprehension test correlations from a classical study of concurrent validity (Bormuth, 1962).

Part of the variation in the cloze-comprehension correlations is almost certainly due to differences in the reliabilities of the measures used in the research. The lower the reliabilities of either the cloze or the comprehension tests, the more attenuated the correlations among these tests. Since the concurrent validity studies have not used equally reliable measures variation in correlations is to be expected, and, therefore, some amount of that variation is explained by differences in the reliabilities.

Another possible source of variation across studies is sample differences. Cloze and comprehension measures might be more correlated for some groups of subjects than for others. The use of subjects drawn from different grade levels, for example, could lead to differences in correlations. Again, the contro' of such differences should limit the variation in the reported correlations.

A third possible source of variation in the correlations is imposed by the limited construct validity of the cloze procedure. Comprehension test questions can be written so as to require the use of information from either within- or across-sentence boundaries (Bormuth, Carr, Manning, & Pearson, 1970). Given that cloze is largely insensitive to the use of cross-sentence information, it seems reasonable to expect cloze to be best correlated with tests consisting of large proportions of within-sentence items. Because the between-within sentence distinction was not made in comprehension tests until recently, the proportions of within- and across-sentence items were uncontrolled. Thus, the reading tests used in concurrent validity studies may not have been tapping the same sorts of comprehension measured by cloze.

To examine some of these issues we attempted to re-analyze the data from the influential Bormuth (1962) study. This data set is reasonable to use to explore the issues discussed here, because it represents one of the most ambitious attempts to demonstrate the concurrent validity of cloze. Bormuth collected cloze test and comprehension test data from a reasonably large sample of subjects at three different grade levels; he reported reliabilities for the various measures used in the study; he made available the cloze passages and comprehension test questions; and, he reported more than 200 separate correlation coefficients of cloze with comprehension. Specifically, this study re-examined the Bormuth data to determine whether we can account for the concurrent validity of cloze in terms of the types of questions used to assess comprehension. Of interest is the within-sentence question variable drawn from the construct validity literature. The variables of reliability and grade level were included also.

Method

This is an archival study. It provides additional analyses of historically important data. These data, because they were collected by a different researcher (Bormuth, 1962), for different purposes from those of the present authors, impose certain limitations upon the analyses. Nevertheless, it is the authors' contention that it is reasonable to reanalyze the Bormuth data so as to determine whether a theoretical construct (e.g., intersentential information integration) can explain extant research results. This reanalysis is not intended as a criticism of the Bormuth research, but rather it is an attempt to extend and explain his findings in a theory-relevant manner.

Bormuth's Sample

Bormuth collected complete data on three grade level samples of 50 subjects. These subjects were attending grades four, five, and six in schools in Indiana.

Bormuth's Materials

Nine passages of 275-300 words in length were written for use in this study. These passages covered three different subject matters (stories, history, science) and they were written at three levels of readability (fourth, fifth, sixth). These passages were transformed into 5th-word deletion cloze tests. Multiple-choice comprehension questions were written for each passage to measure comprehension of vocabulary (12 items per test); factual recall (7 items); sequential order (2 items); cause and effect (4 items); inference (4 items); and author's purpose (1 item). Each of the 31 test items consisted of a sentence stem and four multiple-choice sentence completion items.

Bormuth's Procedures

All subjects completed the nine cloze tests and the nine comprehension tests. Subjects completed the cloze test for a passage before receiving the corresponding intact passage with the questions. Reliabilities of the tests were determined using the rational equivalence method. Correlations were reported between the comprehension tests and the cloze tests. This was done separately for each passage at each grade level, and for each question type across passages at each grade level. Additional information on the Bormuth data is available in the original report.

Reanalysis

Two separate reanalyses of the Bormuth data were completed. In order to reanalyze the Bormuth data it was necessary to identify the number of comprehension test items requiring within- and across-sentence information. To accomplish this a set of rules was drafted based on the system reported by Bormuth, Carr, Manning and Pearson (1970). Separately, the present researchers applied these rules to all nine tests. Inter-rater reliability for the set of nine passages was .87, although some items (vocabulary, sequence, main idea) were rated more consistently than were others (inference, factual). For example, the reliability of the sequential items was 1.00, while it was only .75 for the inference questions. Disagreements were resolved through discussion.

The first reanalysis used the correlations of cloze with comprehension as a dependent measure. Bormuth reported correlations for each cloze passage and

its corresponding comprehension test for each grade level. In all, 27 such correlations were reported ranging from .46 to .87. Log linear transformations of the correlations were performed to put them in an appropriate form for analysis. The log transformations of these correlations were analyzed through multiple regression analysis. The correlations were regressed on the reliabilities of each of the tests, the grade levels of the subjects, and the proportion of within-sentence questions.

A second analysis used the correlations of cloze with each of the Bormuth question categories (i.e., vocabulary, factual recall, etc.) for each grade level. The 189 correlations ranged from .11 to .92. The log transformations of these correlations were regressed on the reliabilities of the two measures, Bormuth's question types categories (a categorical variable indicating which of the passages and which question types were being used), grade level, and proportion of within-sentence questions. The Bormuth question type variable was added to this analysis because the reliabilities for question types were calculated across all nine passages, rather than for each separately. Thus, the reliability coefficients provided are only an approximation of the reliabilities of a particular set of questions. Adding this categorical variable to the analysis should reduce the error introduced by these reliability estimates.

Results

The results of the first regression indicate that the reliability variables combine to explain a significant proportion of the variation in the cloze-comprehension test correlations ($R^2 = .39$, $F\ (2,24) = 7.51$, $p<.01$). The

grade level variable accounts for an additonal 24 percent of the variance (F (1,23) = 14.41, $p<.005$). Finally, the proportion of within-sentence questions accounts for 8 percent of the correlation variance with the other independent variables already in the regression (F (1,22) = 4.94, $p<.01$).

The second analysis, the regression of the correlations of scores on the various Bormuth question types with cloze scores, again found the reliabilities to account for a substantial amount of variance (R^2 = .66, F (2,186) = 185.10, $p<.0001$). Grade level accounts for an additional 6 percent of the variance (F (1,185) = 38.27, $p<.0001$). The Bormuth question type variable was also significant (F (2,183) = 6.36, $p<.05$), although it explained only 2 percent of the variance. Proportion of within-sentence questions was then allowed to enter the regression. This variable accounted for less than 1 percent of the variance (F (1,182) = 2.56, N.S.). The within-sentence question variable was significantly related to the cloze-comprehension correlations (r = .19, $p<.05$), but its contribution dropped to almost nothing with the other variables in the equation.

Discussion and Conclusions

This study attempted to account for the variation in the correlations between cloze scores and comprehension test scores. The results of reanalysis of the Bormuth data indicate that reliability and sampling differences accounted for some of the variation in the correlations. However, these variables were not sufficient to account for all of the explainable variation. For this reason, the proportion of comprehension test questions found to assess within-sentence information was then used as an independent variable.

Nevertheless, the results from the regression analyses were mixed. The within-sentence variable increased the explanation of cloze-comprehension correlations significantly in one analysis, but not in the other.

The differences in the findings could be due to a number of factors. Reliabilities of the tests used in the two reanalyses were measured more directly in one analysis than in the other. That is, in the first reanalysis they were based on the total comprehension test scores. In the second, they were based on the Bormuth question type scores, but the reliabilities for question types were computed across tests while the correlations were reported within tests. To compensate for this problem a Bormuth question-types variable was used. This variable was significantly correlated to the within-across variable under study. Also, the within-across distinction was made more reliably for some question types than for others. The error introduced by this was balanced across tests in the first reanalysis, but it was separated and unequal in the other. Such differences could have led to different findings.

The findings suggest that the limited construct validity of cloze as a measure of comprehension has a small but significant influence on the concurrent validity of cloze with measures of comprehension. Cloze predicts performance best in those situations in which within-sentence comprehension dominates. Cloze is less useful in situations in which students are expected to integrate information across sentence boundaries.

Experimental research is now needed to explore these findings. Such research could improve upon the design used here in a number of important

ways. First, the order of test administration should be balanced, in order to reduce spurious increases in the correlations attributable to order effects. Second, questions could be designed specifically to assess within- and across-sentence information. Third, prior knowledge assessment should be made to determine the text dependence of the questions. If a question is answerable on the basis of prior knowledge, then it is neither a within- nor an across-sentence item. Fourth, question-answer items could be designed to replace Bormuth's cloze-like sentence completion items. This would permit a more accurate estimate of the cloze-comprehension correlations. Fifth, reliabilities could be measured more accurately using Cronbach's alpha. This would reduce spurious increases in the relationship of reliability with the cloze-comprehension correlations. A study which deals with these problems will be better able to account for the reasons that cloze is not always well related to comprehension.

Until such a direct test of the influence of construct validity upon concurrent validity is made, two cautions seem reasonable. Those who use, or recommend the use of, cloze as a useful instructional placement device should beware of the great amount of variation in the relationship of cloze with comprehension measures. This suggests that cloze can be, <u>but is often not,</u> a useful prediction device. Also, this analysis indicates that cloze is, under some conditions, better related to within-sentence comprehension demands than to across-sentence ones. If across sentence comprehension is important, then there is a possiblility that cloze will be somewhat less effective as a predictor.

References

Bickley, A.; Ellington, B.; & Bickley, R. The cloze procedure: A conspectus. Journal of Reading Behavior, 1970, 2, 232-249.

Bormuth, J.R. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation, Indiana University, 1962.

Bormuth, J.R.; Carr, J.; Manning J.; & Pearson, P.D. Children's comprehension of between- and within-sentence syntactic structures. Journal of Educational Psychology, 1970, 61, 349-357.

Kibby, M. Intersentential processes in reading comprehension. Journal of Reading Behavior, 1980, 12, 299-312.

McKenna, M.C.; & Robinson, R.D. An introduction to the cloze procedure. Newark, Delaware: International Reading Association, 1980.

Rankin, E. The cloze procedure--a survey of research. In E. Thurston, & L. Hafner (Eds.), The philosophical and sociological bases of reading. Fourteenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: National Reading Conference, 1965.

Shanahan, T.; & Kamil, M.L. The sensitivity of cloze to passage organization. In J. Niles, & L. Harris (Eds.), New inquiries in reading research and instruction. Thirty-first Yearbook of the National Reading Conference. Rochester, New York: National Reading Conference, 1982.

Shanahan, T.; & Kamil, M.L. A further comparison of sensitivity of cloze and recall to passage organization. In J. Niles, & L. Harris (Eds.), Thirty-second Yearbook of the National Reading Conference. Rochester, New York: National Reading Conference, in press.

Shanahan, T.; Kamil, M.L.; Tobin, A.W. Cloze as a measure of intersentential comprehension. Reading Research Quarterly, 1982, 17, 229-255.